White Paper Report

Report ID: 98496

Application Number: HD5057409

Project Director: Martin Foys (mfoys@drew.edu)

Institution: Drew University

Reporting Period: 5/1/2009-5/31/2010

Report Due: 11/30/2010

Date Submitted: 12/19/2010

December 15, 2010

Final Report for the Digital Mappaemundi Project

Recipient of an NEH Digital Humanities Startup Grant: May 2009 – September 2010

Co-directors:

- Martin Foys, Department of English, Drew University (<u>mfoys@drew.edu</u>)
- Shannon Bradshaw, Department of Computer Science, Drew University (sbradsha@drew.edu)
- Asa Mittman, Department of Art and Art History, California State University, Chico (asmittman@csuchico.edu)

1. Originally Stated Goals (from initial 2009 application to the NEH)

Long term: "Digital Mappaemundi (DM) proposes an extensible, collaborative model of open source scholarship, one that allows scholars to collaboratively contribute to and edit associated networks of image and text data, and users to search and organize this material in ways impossible before. As such, this work should prove applicable and valuable far beyond its initial scope of study. To develop such a new architecture, this project will focus on medieval mappaemundi ("maps of the world") and the geographic texts on which they were based. The transdisciplinary nature of these maps and texts provide (following the argument of the ACLS Cyberinfrastructure Report [see < http://www.acls.org/programs/Default.aspx?id=644>]) a perfect basis for the development of a cooperative and open source tool well-suited for adaptation and use for other humanities projects."

Short term (life of the grant): "Currently, our most significant funding need is the refinement of our XML schema, completion of the database, and development of the front-end software application for *Digital Mappaemundi*. With institutional seed funding, we have established the bare bones of these systems, which now need considerable expansion and modification. We need to employ a programming team of student assistants, who will finish the database architecture and develop the front-end GUI interface. With the aid of this team, we also need to refine the syntax, lexicon and functionality of our XML tagging and database mapping, based on feedback from the initial prototype. To further refine this work, we need to conduct evaluations of the project."

2. Activities

Work on *DM* proceeded on two fronts: the generation of analyzable data from medieval maps and geographic texts, and the development of the architecture to make such generation and subsequent analysis possible.

A. *Data Generation*: Over the course of the grant term, Martin Foys, Asa Mittman and a team of graduate and undergraduate student research assistants used an early prototype of the *DM* editing feature to annotate seven medieval maps and

two geographic texts – the foundation for a small but representative sandbox of data. All textual inscriptions and descriptions of locations on maps and in texts were recorded (with coordinate locations for images and word placement for texts), along with accompanying descriptive data. This dataset, while relatively modest in size, provided excellent material through which to work out major issues of design and functionality. This process resulted in major shifts and improvements in the vision for this project. In addition, we now have a robust dataset on which to build and through which to test our interface as the resource develops. The dataset may also serve as a framework for subsequent users to model.

Material edited/annotated

(Note: much of this material remains the domain of various institutions, and so is not currently freely available to the public; starting in Spring, 2011, scholars will be permitted access for continuing testing and development, on a case-by-case basis. See p. 8, below for further information.)

Maps:

- the Anglo-Saxon Cotton Map (British Library MS Cotton Tiberius B.V. f. 56v)
- a Higden *Polychronicon* Map (Corpus Christi College Cambridge [Parker Library], MS 21, f. 9r)
- the Thorney Map (Oxford- St. John's College MS 17, f. 6r); the Sawley Map (Corpus Christi College Cambridge [Parker Library], MS 66, p.2)
- Matthew of Paris's *Chronica Maiora* Map of the Middle East (Corpus Christi College Cambridge [Parker Library], MS 26, f. 3v-4r)
- an unfinished late Anglo-Saxon *mappamundi* (Corpus Christi College Cambridge [Parker Library]; MS 265, p.210
- a *mappamundi* from Gunther Zainer's 1472 printing of Isidore's *Etymologiae* at section III ("De Asia") of Book XIV ("De Terra")

Texts:

- Book One, Chapter Two of Orosius's *Adversus Paganos Historiarum Libri Septem* ("Seven Books Against the Pagans")
- Book One, Chapter One of the Old English Translation of Orosius's *Adversus Paganos Historiarum Libri Septem* ("Seven Books Against the Pagans")
- B. *Design and Functionality Development*: At the same time that teams under Mittman and Foys were generating core data, a programming team under Shannon Bradshaw was developing early prototypes of the editing and annotation resources. The earliest pre-alpha prototype was implemented in Adobe Flash and employed single coordinate-point visual tagging. Annotations were constrained to a

rudimentary set of XML descriptor tags and keywords. See Appendix A, figures A.1-A.4 for screenshots of this phase 1 prototype.

After a few months of development along these lines, our thinking progressed as we gradually understood the limitations inherent in such a traditionally framed approach. One of the watershed moments in the project's development was the evaluative summit the project held in October 2009, where a diverse group of specialists (e.g. map specialists, new media scholars, college professors, metadata librarians) were brought together at Drew University for a weekend to try out and respond to the phase 1 iteration of the resource. The feedback and "blue sky" discussions that followed made it clear that we should retool our approach to designing such a resource.

Subsequently, Foys, working with Bradshaw, developed a new architecture for visual tagging, based on a tripartite approach of "Dot-Line-Shape" (D-LISH). D-LISH was designed to provide maximum flexibility with regards to annotators' graphic desires, providing them with the ability to more generally indicate annotated areas on a map through single coordinate points, joining points of significance lineally (either through traditional lines of text, or by connecting horizontally disparate points [i.e. by "crooked lines"]), or, most granularly, by demarcating areas of annotation within user-generated polygonal spaces. Further, a single annotation could then be grouped with any combination of multiple dots, lines and/or shapes. The result was an image-selection tool that allowed for considerably greater choice in determining what, exactly, was to be selected and annotated on an image than in pre-existing tools for digital scholarship.

At the same time that D-LISH was being developed, the DM team was re-thinking its approach to annotation and, by relation, search. At the beginning of our work, we approached annotation through traditional modes of an XML-dictated philosophy that imagined discrete sets of tags that "covered" the taxonomy of content medieval maps and related texts might contain (e.g. a field for inscription, translation, keyword, figure, continent, quality categories [city, people, river, mountain, etc.], and so forth), and a search function that generated results by polling one or more of these selected fields. Such a taxonomic approach is perhaps most readily understood as analogous to the patterns of data annotation demanded by such protocols as TEI (Textual Encoding Initiative), which imagines a prescriptive set of tag fields that can comprehensively cover, in theory, any aspect of a text being marked up in XML which might need description. We were perhaps lucky, in that that the content of medieval maps is so heterogeneous and unpredictable that relatively early in the sandbox phase of our development, we realized the considerable limitations of such a taxonomic approach. We really had no way of adequately first imagining then accommodating the ways in which future users might wish to annotate the materials we were editing, or how we might keep our own prescriptive categories for annotation free from our own heuristic bias.

In response, the *DM* team radically shifted the direction of development with respect to models of annotation and search, moving towards a completely open mode of annotation, where users simply enter annotations as untagged text, with the option to designate individual terms as prioritized keywords, through the use of a hash (#) tag. Such annotation cannot really be considered purely folksonomic, though it does retain the capacity for such tagging if desired by the user. In response to such a simpler, open-form annotation model, the development of our search algorithm has become substantially more complicated, as the work of parsing annotated material will have to be handled by the search, not the annotation function. In the end, however, this model will result in a substantially more agile and dynamic response for end users wishing to search a growing set of both materials and annotation. Much more work will still need to be done in this area.

In addition, we also began moving towards the development of a collaborative model for the addition to and editing of DM materials, where users can upload their own materials for annotation, designate private or public layers of annotation of materials, and create multiple layers of differentiated annotation, or collaborate with specific users on the annotation of one or more layers. Searching, as well, can then happen in limited modes not across tags (as in the traditional XML-driven model) but instead delimited within or across specific layers of annotations, and can then result in dynamically constructed layers built out of results returned for any given document. Envision, for instance, in a search on a map for specific authorial sources (e.g. Orosius, Jerome, Isidore, Aethicus Ister) a graphic layer returned with all annotated areas containing such names, which could then be manipulated to show particular sources individually or comparatively (by color or color gradation), or then further expanded or delimited by other variables. Such an environment is still very much in the planning stages, but the work completed in the above described areas has laid the foundation for this next phase of development. The ability to return dynamically and graphically the results of searches for image-based annotation in this fashion would represent a significant step forward in the way in such scholarly work may be done digitally.

See also appendix A, figures A.5-A.8 for representative screenshots of the phase 2 prototype's features, including D-LISH. We are currently planning to have a alpha rollout of this resource in early Spring 2011, though access to some material will be contingent upon discussions with the institutions which hold these materials. See page 8, below, for further information regarding this release.

Accomplishments

The section on "Activities," above, already details many of the specific accomplishments of term of the start-up grant funding. The most important point that needs to be made in this section is how important it is that we *did not* fulfill the majority of our goals as described in our application. Our original plan of refining and augmenting the

architecture of traditional XML tagging and taxonomy very quickly became outmoded by the series of conceptual breakthroughs our team had during the first six months of our work, as we labored to fulfill this self-imposed mandate.

Instead, in addition to the digital editing and annotation of a robust starting dataset (see above), the *DM* team implemented a groundbreaking system of image area selection (D-LISH, above), and laid the foundation for further advances with regards to open-form editing, annotation and searching (see also above). In addition, Bradshaw's team completely redesigned and reprogrammed the phase 1 resource, which had proved useful for quickly working out several design and functional issues. In phase 2, the user interface and the backend data model, annotation repository, and search systems were completely rebuilt based on lessons learned in phase 1. In phase 2, our objective is a broadly applicable annotation toolbox and repository.

In carrying out this phase, the backend database was redesigned and rebuilt with a more general data model. This model enables us to easily accommodate data in the Open Annotation Collaboration (OAC: http://www.openannotation.org/) protocol, which is emerging as an interoperability standard among large repository and tools initiatives underway at Stanford University, Johns Hopkins University, the University of Toronto, and the University of Maryland, among others. As a second step toward a broadly usable resource, we have exposed *DM's* annotation repository through a web services framework. The web service enables client systems (including our own) to search our annotation repository using an expressive query language and sophisticated search metrics. The search system is based on Solr from Apache's Lucene project. Clients may search the repository using raw search terms, faceted search by user-defined field, or more complex techniques such as more-like-this queries. Client systems may also mine or store annotation data by pulling or pushing XML feeds of layered annotations for any map in our dataset. Clients may also constrain annotation data feeds by proximity to a location on an image, by annotating user, by timestamp, and a variety of other metrics. Additionally, external URI links to specific coordinates and annotations may be generated.

In rebuilding the front end, we moved from a Flash-based architecture to one that is based on Javascript. In phase 2 development of the user interface, we have been careful to develop toward an open-source library that enables widespread adoption. We now have an alpha version of a set of tools to enable scholars to mark regions-of-interest within images and associate textual annotations with those regions. Scholars may mark images with individual points, segmented lines, or custom polygonal shapes. Significantly, a scholar may identify any number of markers on any number of images as the targets for textual annotation. Additionally, a given marker may serve as the target for any number of textual annotations. Finally, scholars may organize their annotations into groups called 'layers' so that different research questions involving a single image may be addressed separately through annotation. Scholars may choose to view a single layer of annotation or view multiple layers of annotation overlayed on one another. Our current toolbox may be integrated with existing tools on other web sites (i.e. manuscript repositories) with

some modification. In Spring 2011, we will publish a comprehensive open-source annotation toolbox that will enable other sites to easily integrate its functionality.

So while many of the initial goals of this project have not been met, they never will be, as these goals have been dramatically re-written. But work achieved in replacement of these goals has been equally as dramatic, and the *DM* team is confident that we have more than surpassed our initial expectations for the work we might achieve.

Audiences/Dissemination/Funding

As this grant was a start-up grant, with the specific goal of in-house development of a resource for future public use, the audience for this resource remains at this point largely potential rather than actual. However, reports of our work have been disseminated through a series of conference and symposium presentations (see below, and Appendix B for representative supporting materials), as well as a recent *New York Times* article on the state of Digital Humanities http://www.nytimes.com/2010/11/17/arts/17digital.htm. At this point, awareness of the *DM* in medieval studies and digital humanities circles is quite high, and we are well positioned to begin recruiting scholarly test users for our alpha rollout in early Spring 2011. This summer, the co-directors will be presenting on this release at both major medievalist conferences (Kalamazoo and Leeds), as well as an international symposium on cartographic studies being held at Oxford University this summer. In addition, *DM* was awarded a modest (\$16,000) scholarly communications grant from the Andrew W. Mellon Foundation to explore a use-case study in interoperability between this resource and Stanford University and Cambridge University's Parker-on-the-Web project (see Appendix B, figures B.7-8).

DM-related public presentations:

Given:

Bradshaw, Shannon. "Annotation Interface Tools in *Digital Mappaemundi*." Stanford University, Andrew W. Mellon Foundation Digital Manuscript Studies Technical Meeting, #2, 10 December 2010.

Bradshaw, Shannon, and Foys, Martin. "The *Digital Mappaemundi* Resource" MARGOT Conference, Barnard College, New York, NY, 28 June 2010.

Mittman, Asa. "Navigating the Margins: Sources, Analogs, Wandering Monsters and the *Digital Mappaemundi*," the International Medieval Congress, University of Leeds, July 2010.

Foys, Martin. "Medieval Maps: Datascapes and the Next Generation of Digital Scholarship." University of Illinois, Champaign-Urbana, IL, 2 April 2010.

Foys, Martin. "Interoperability, Intermediate Data, and Scholarly Siloing in Current Digital Manuscript Resources." Medieval Academy of America, Yale University, New Haven, CT, 20 March 2010.

Mittman, Asa. "Where have all the monsters gone?: The Monsterless Maps of Matthew Paris" Andrew W. Mellon Foundation Scholarly Communications Symposium, Paris, France, 15 January 2010.

Bradshaw, Shannon and Foys, Martin. "Parker-on-the-Web and *Digital Mappaemundi*: a Use-Case Study in Interoperability." Andrew W. Mellon Foundation Scholarly Communications Symposium, Paris, France, 15 January 2010.

Foys, Martin. "The *Digital Mappaemundi* Project: Connections, Collaboration, Consistency, Community." Delaware Valley Medieval Association (Keynote), University of Pennsylvania, Philadelphia, PA, 4 December 2009.

Upcoming:

Foys, Martin and Mittman, Asa. Panel Discussants, "Proposing a Toolkit for Medievalists" (accepted) International Congress on Medieval Studies, Kalamazoo, May 2011.

Foys, Martin. "Unarchiving the Map" (proposed). *The Language of Maps: Communicating through Cartography during the Middle Ages and Renaissance*. Oxford University, June 2011.

Bradshaw, Shannon, and Foys, Martin. "Agile Cartography" (accepted). In *Mappings: Digital Cartographies*. International Medieval Conference (IMC), University of Leeds, England, July 2011.

Mittman, Asa. "Navigating Multiple Worlds" (accepted). In *Mappings: Digital Cartographies*. International Medieval Conference (IMC), University of Leeds, England, July 2011.

Publications:

Foys, Martin and Mittman, Asa. "Digital Mappaemundi: Changing the Way We Work with Medieval World Maps." Peregrinations Journal of Medieval Art & Architecture (2.3: 2009), 170-177.

Bradshaw, Shannon, and Foys, Martin. "The *Digital Mappaemundi* Resource" (submitted). *Selected Proceedings of the "Third International MARGOT Conference, The Digital Middle Ages: Teaching and Research."* Ed. Christine McWebb and Helen Swift (Digital Medievalist: 2011).

Continuation and Long-Term Impact

This funding – and the work and invitations and networking it led to – has been invaluable for making connections with other partners (Stanford University, the British Library, University of Cambridge, the Bodleian Library (Oxford), Kings College London, University of Toronto) and opening up possibilities of additional funding to continue the

project's development. Through the scholarly and public discourse noted in the section above, the project has already identified and had communication with a dozen test-users ready to start using and/or contributing to (and in two cases, funding further) work in *DM*. And, after the *New York Times* article three weeks ago, we set up an access list for next year's alpha rollout; it already has over fifty users subscribed. Finally, and perhaps most significantly, our work developing the D-LISH system of image-area selection has pointed the way to further collaboration with Parker-on-the-Web (http://parkerweb.stanford.edu/) and other long-term projects in the digital humanities, including the new and massive Open Annotation Collaboration (OAC) resource. The next several months will be spent showcasing the work we've done so far, selectively signing on partners for robust use-case testing (through the funded and/or volunteer development of materials both cartographic and non-cartographic in nature), and the procurement of funding for the next cycle of development. Additionally plans are underway to apply for a higher level of Mellon Foundation funding, and, with UK partners, to apply for the next round of NEH "Digging into the Data" grants.

For the short term, we will continue our planned rollout of the phase 2 alpha prototype this coming Spring, at which point we will invite other scholars to begin using and contributing to *DM's* primary source material and annotations for testing and development purposes. Mittman has an institutional grant to port the digital files of Thomas Ohlgren's *Insular and Anglo-Saxon Illuminated Manuscripts: An Iconographic Catalogue c. A.D. 625 to 1100*, a print index and cross-listed description and keyword "database" of 229 codices accounting for nearly all visual iconography in Anglo-Saxon manuscripts. We have also been provided with a grant from a private donor to use the *DM* annotation system to create a digital edition of a thirty-foot long medieval genealogical scroll (see Appendix B, figure B.6). Such funding will allow further refinement of the *DM* system, while bridging the project to the next cycle of grant applications. The work from these two opportunities will also provide exciting opportunities for proofs of concept of the utility of the annotation system beyond its original focus of medieval maps – a long-term goal of this project, and one stated in our original NEH application.

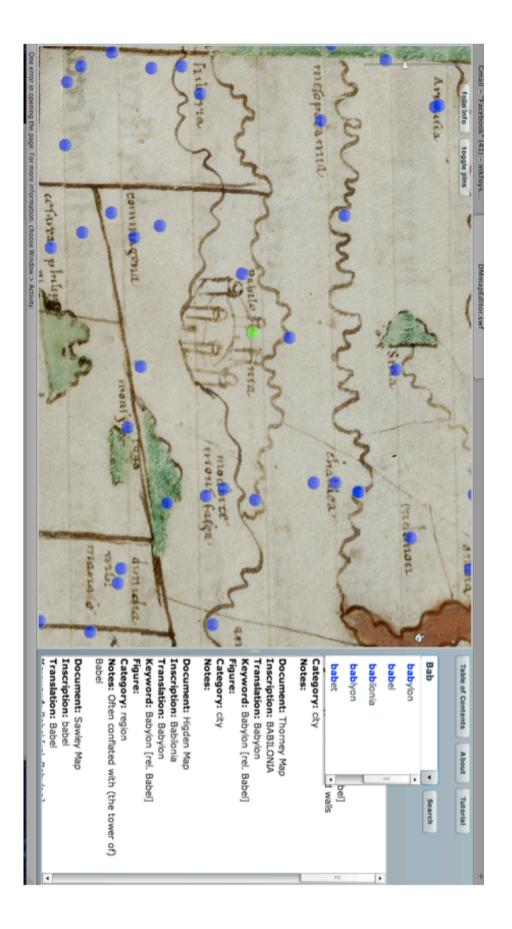
The most exciting long-term prospect (aside from the continuation and growth of the original cartographic focus of the project) is the leveraging of the tools developed for the content enrichment of current digital repositories. Here, the connections we have made through the NEH-funded work to projects and developers at Stanford, Johns Hopkins, University of Toronto, the Andrew W. Mellon Foundation, and, most recently, OAC, have put the *DM* team in a promising position to build upon our start-up development.

Products

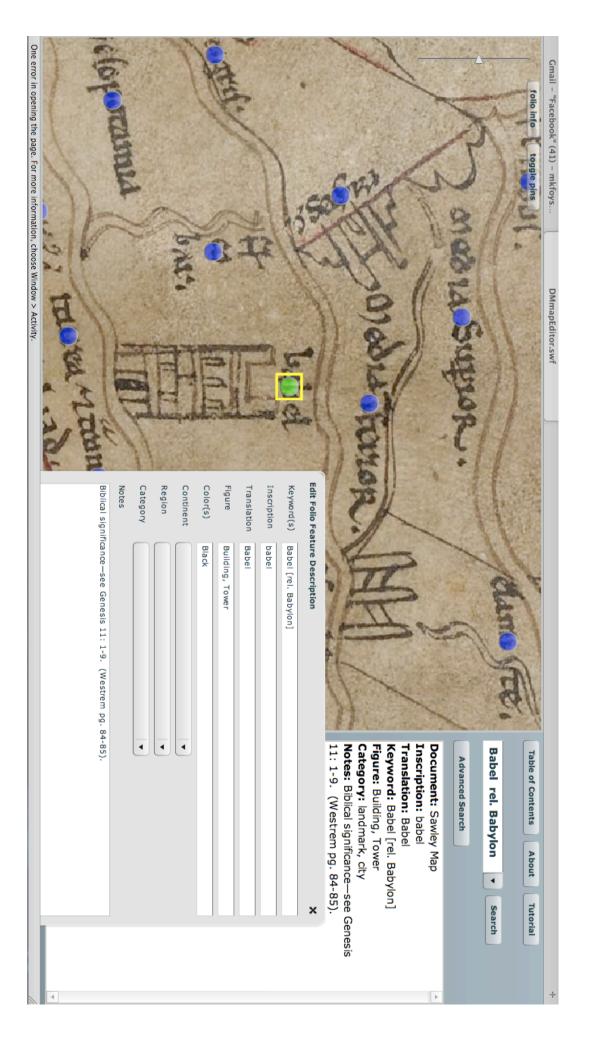
At this time, the phase 2 alpha version is still under development, with an initial release for limited scholarly testing planned for early next year (for more information, visit our site: http://bob.drew.edu/mappaemundi/). For a series of screenshots from both phases of development of DM, see Appendix A, immediately following.

Q http://bob.drew.edu/m Move Annotation Delete Annotation Edit Annotation Continent Inscription Keyword Babel babel Biblical significance—see A
Genesis 11: 1-9. Babel [rel. Babylon] Building, Tower Confirm Edit Figure: city
Continent: Asia
Region: The Holy Land
Category: city
Map: Cotton Map Figure: Continent: Asia Keyword: Babel [rel. Babylon]
Inscription: babel
Translation: Babel Keyword: Babylon [rel. Babel]
Inscription: babilonia
Translation: Babylon Category: city Map: Thorney Map babylor Franslation: Babylon Babel] × D- 6

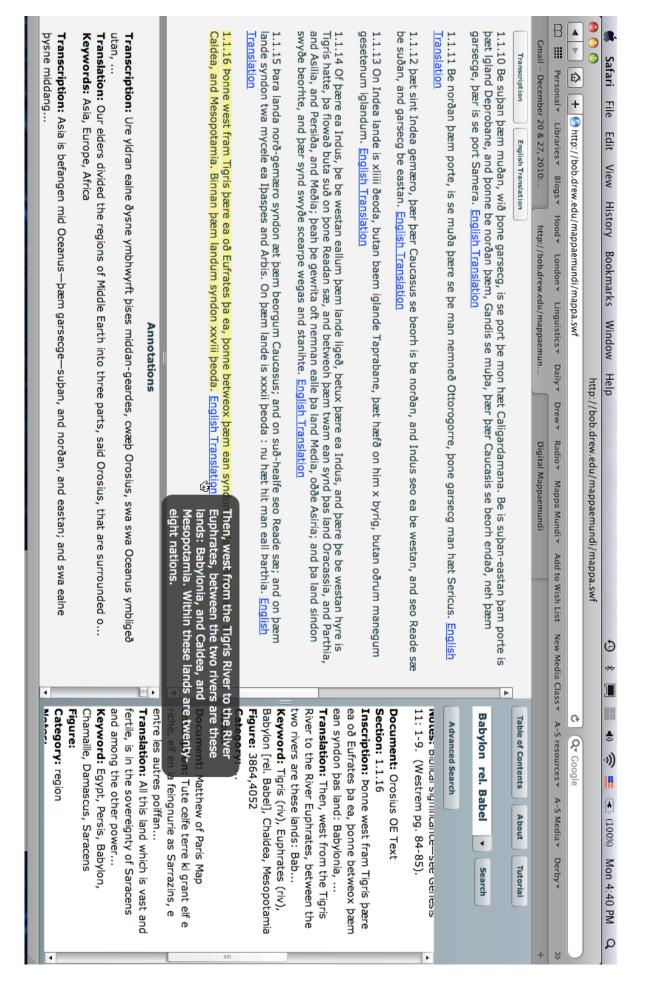
A.1: composite graphic demonstrating many of the annotation features of the phase 1 prototype



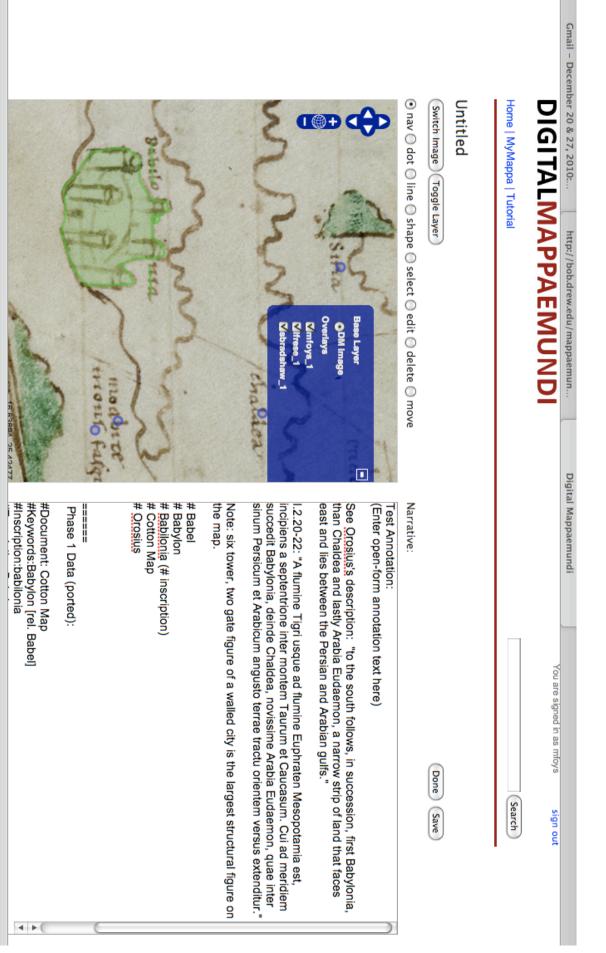
A.2: DM Phase 1 prototype; coordinate editing and incremental searching



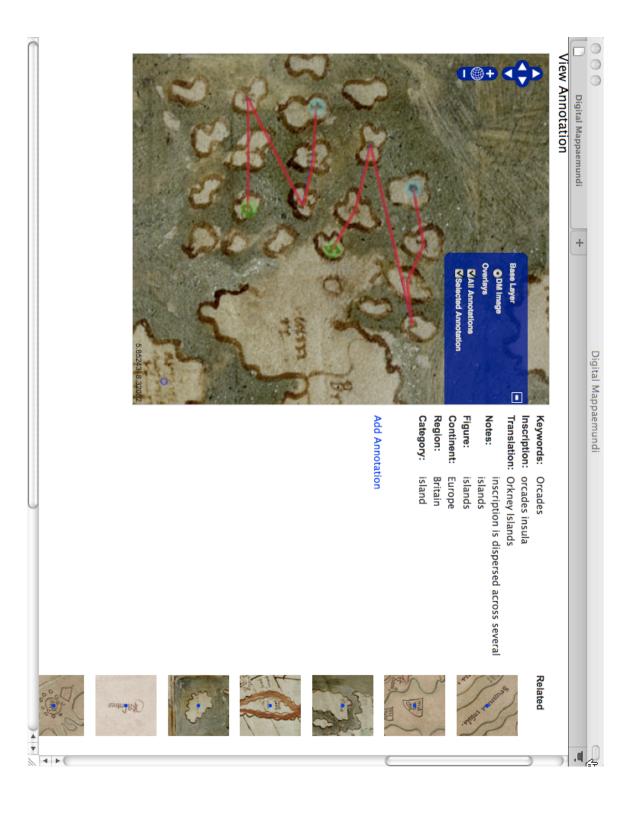
A.3: DM Phase 1 prototype; prescriptive, taxonomic tagging



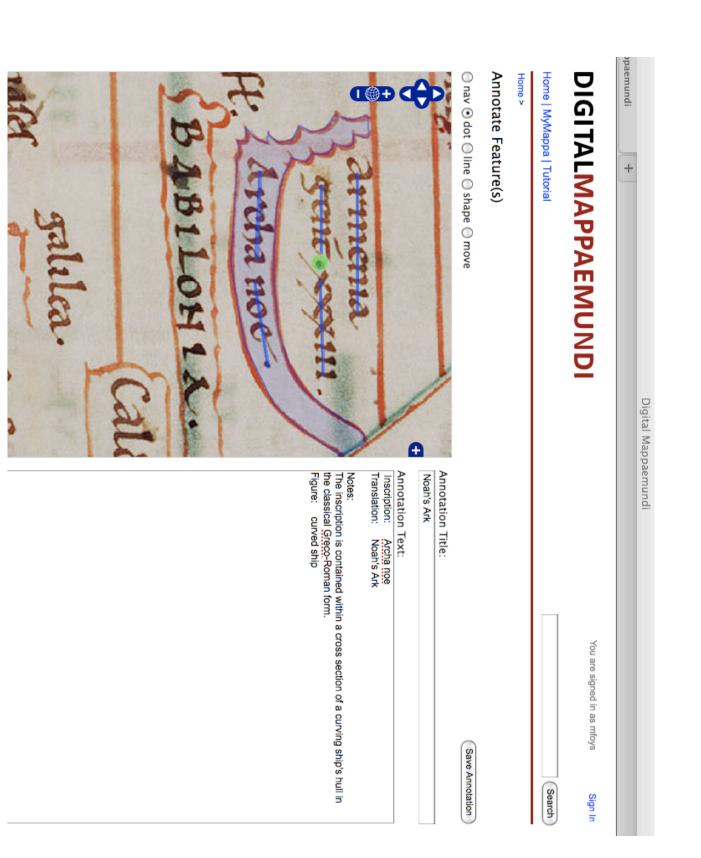
A.4: DM Phase 1 prototype; text display, translation and annotation



user layers, and open-form annotation entry DM Phase 2 prototype; new interface, D-LISH shape, line and dot annotation-area selection,



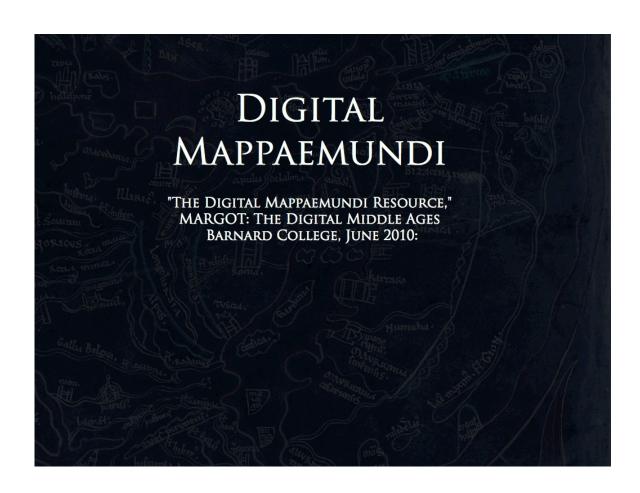
visual analogue feature A.6: DM Phase 2 prototype; D-LISH annotation, segmented line selection for annotation and related

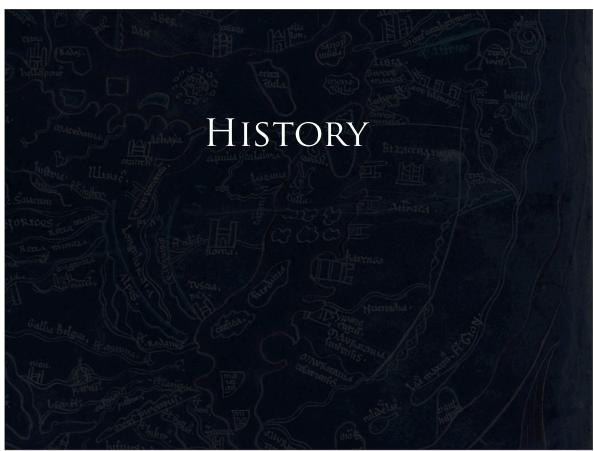


coordinate points in annotation area selection A.7: DM Phase 2 prototype; D-LISH annotation combining polygonal shapes, segmented lines and

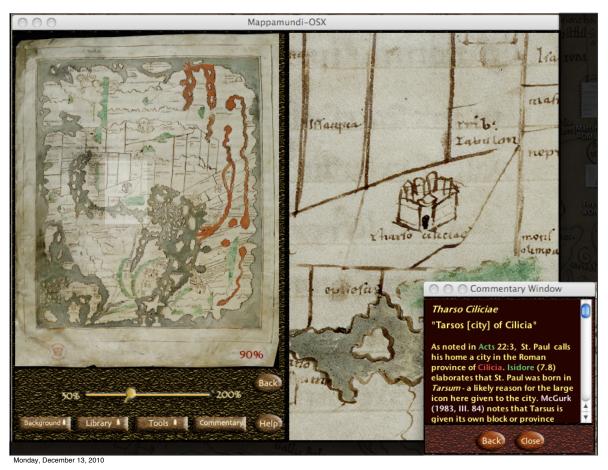
Appendix B:

Representative Keynote deck for *DM*-related presentations, from Bradshaw and Foys: "The *Digital Mappaemundi* Resource," @ MARGOT: The Digital Middle Ages, Barnard College, June 2010:

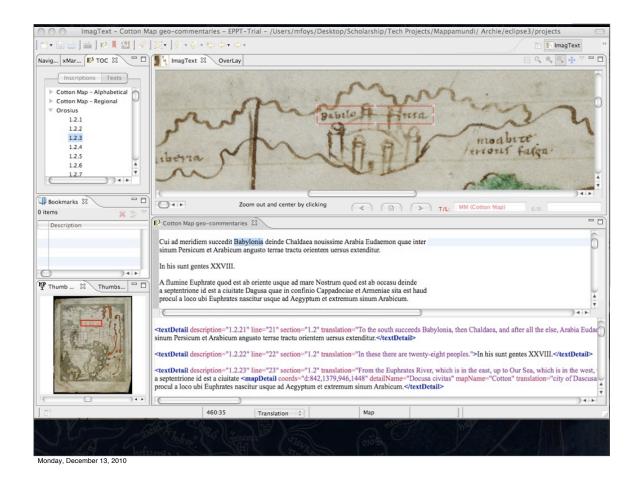


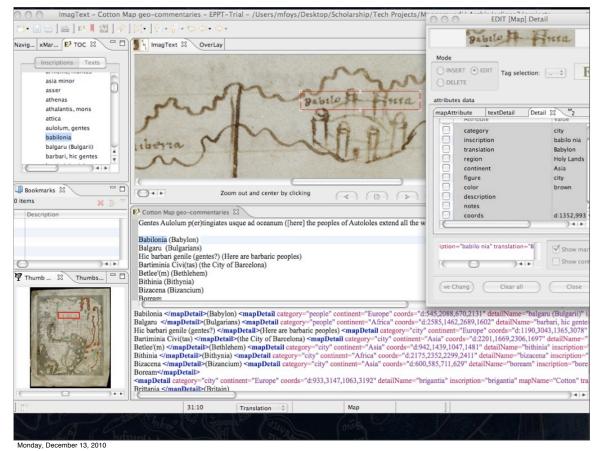


Monday, December 13, 2010

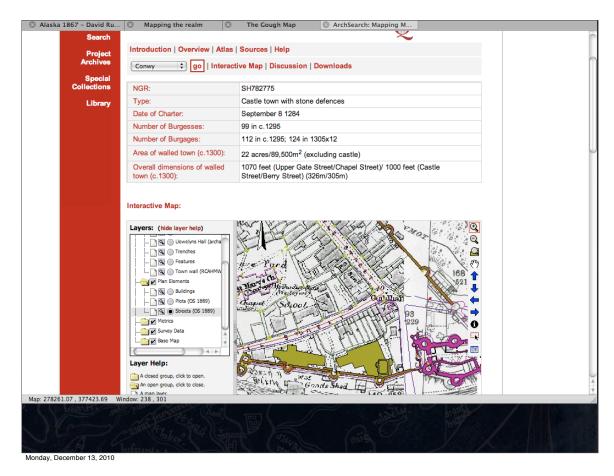


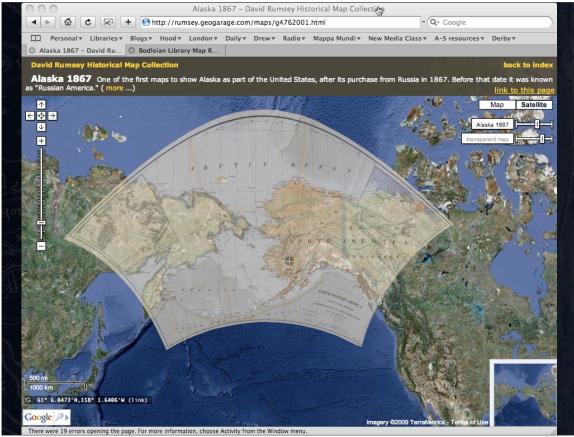
B.1: early attempts at a docucentric edition of the Cotton Anglo-Saxon *mappamundi* (Foys:





B.2: early use of Electronic Presentation and Production (EPPT) software to link coordinate points and texts through XML (Foys: 2005-2006)

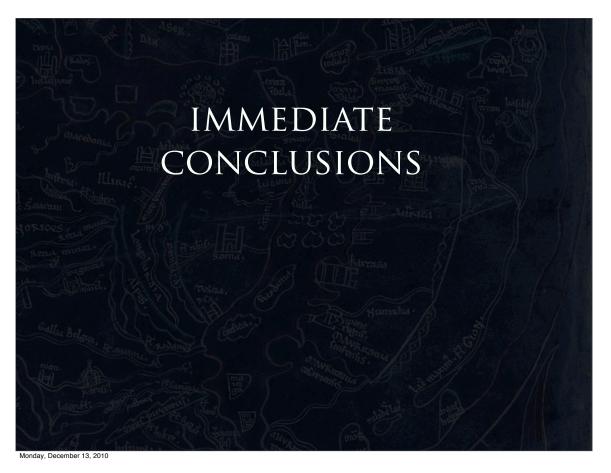


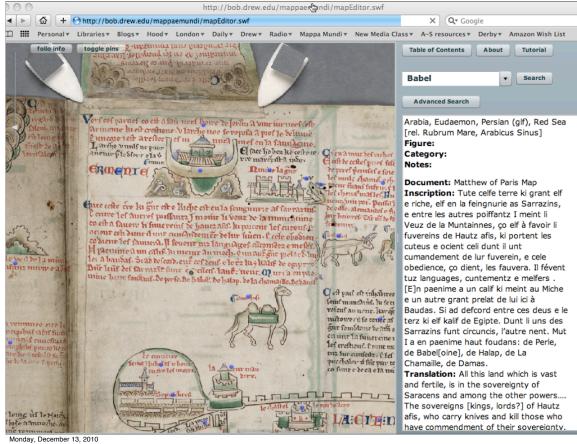


B.3: examples of physical geo-spatial bias in digital approaches to cartography of the past

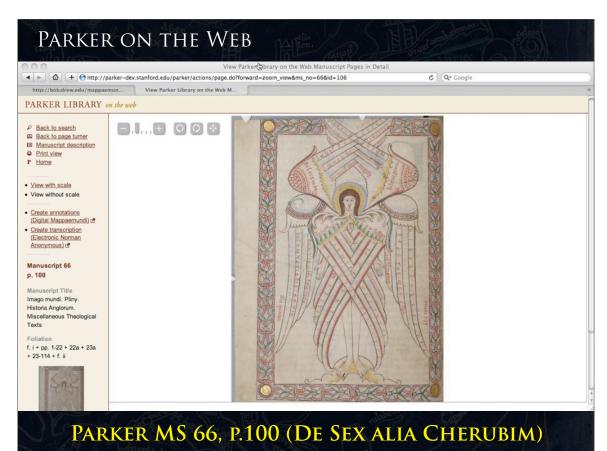


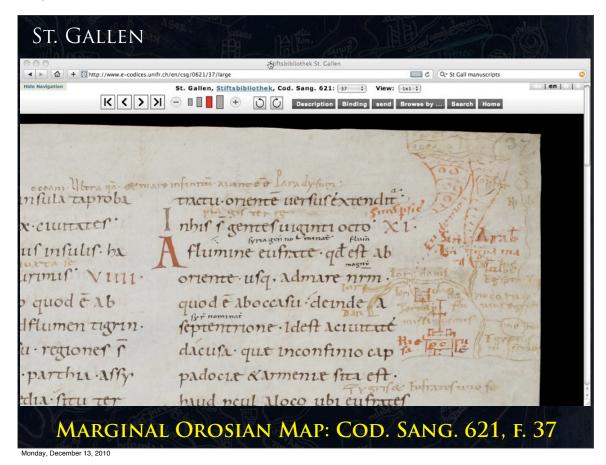
(Series of slides detailing phase 1 prototype of DM originally follow; not reproduced because of general redundancy with screen shots provided in Appendix A.)



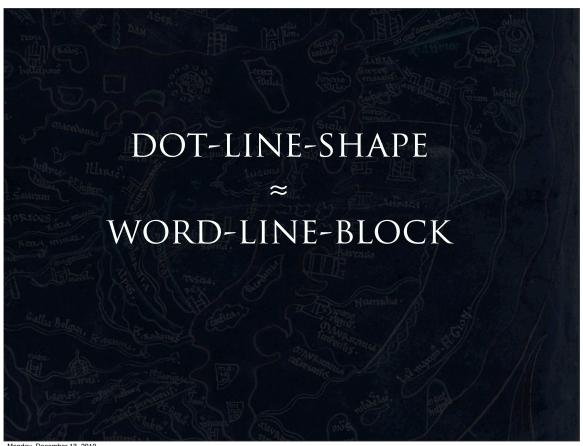


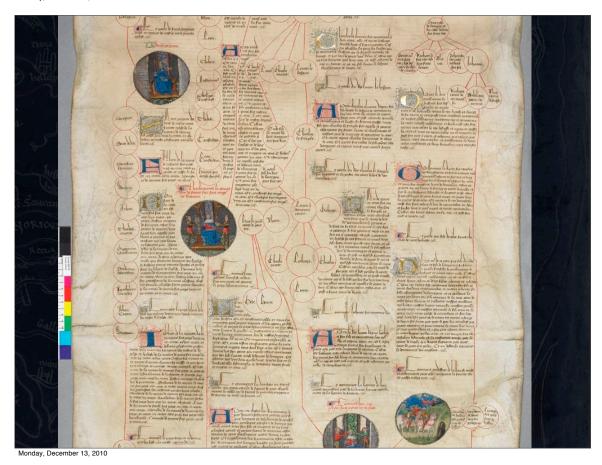
B.4: the limits of coordinate editing: As Mittman's use of *DM* to edit Matthew of Paris's map of the Middle East.





B.5: examples of hybrid-form content in current "siloed" digital manuscript repositories; the challenge of such fluid forms of content





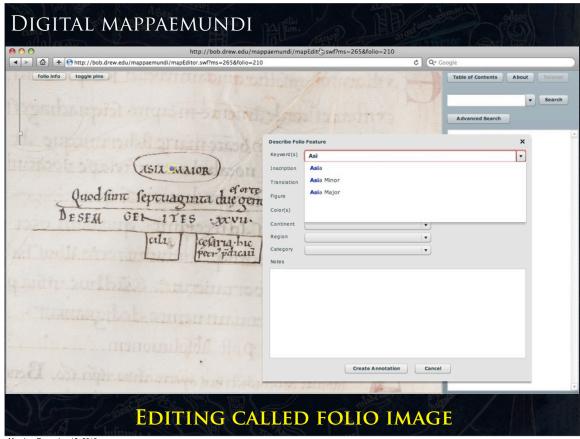
B.6: non-cartographic, funded use case for phase 2 implementation: thirty-foot long medieval genealogical scroll

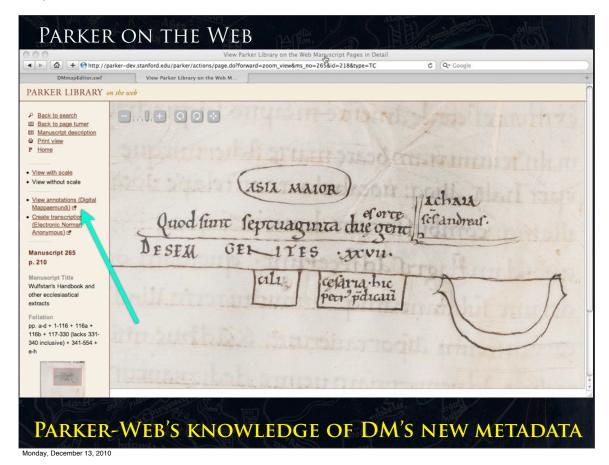


PARKER ON THE WEB View Parker Librery on the Web Manuscript Pages in Detail (4) D (http://parker-dev.stanford.edu/parker/actions/page.do?forward=zoom_view&ms_no=265&id=218&type=TC 9) ☆▼ · G · Google -J...+ 00 · P Back to search Back to page turner ■ Manuscript description
 ● Print view P Home » View with scale » View without scale » Create annotations (Digital Mappaemundi) » Create transcription (Electronic Norman Manuscript 265 p. 210 Wulfstan's Handbook and pp. a-d + 1-116 + 116a + 116b + 117-330 (lacks 331-340 inclusive) + 341-554 + e-h PW: PINGS DM'S WEB SERVER & SEES NO METADATA EXISTS; LINK OPENS FOLIO IN DM RESOURCE

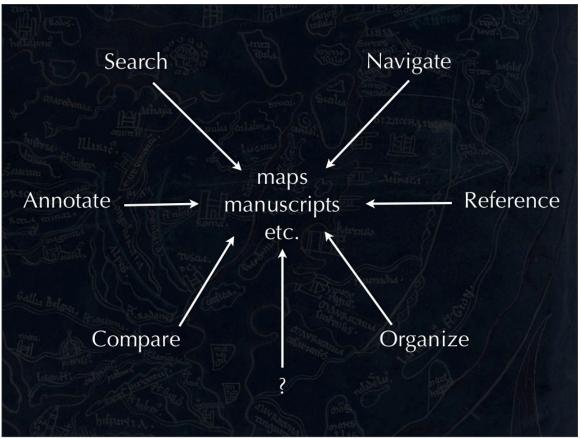
B.7: demonstration of interoperability proof of concept between Phase 1 *DM* and Parker-on-the-Web project (http://parkerweb.stanford.edu/); Parker resource can call out to *DM* and note that image has not been edited.

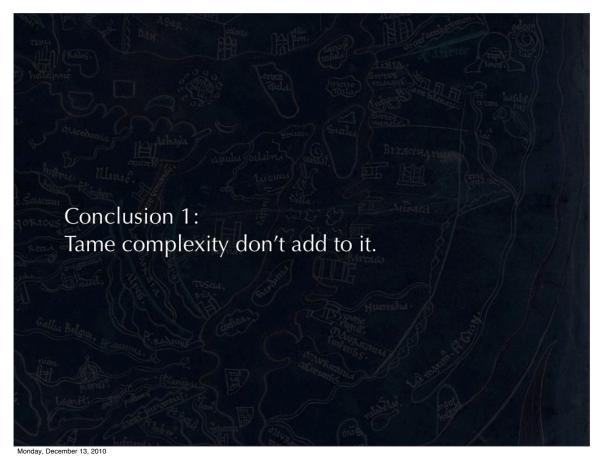
Monday, December 13, 2010



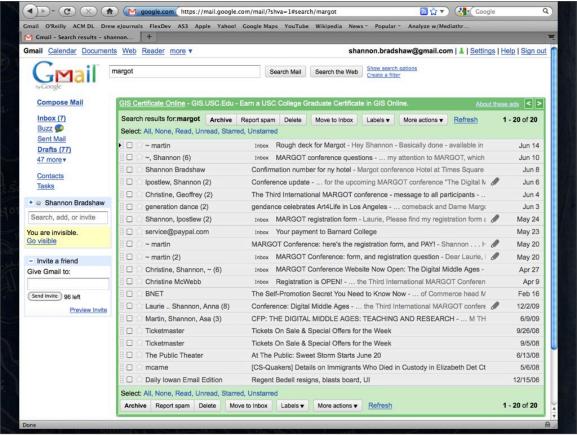


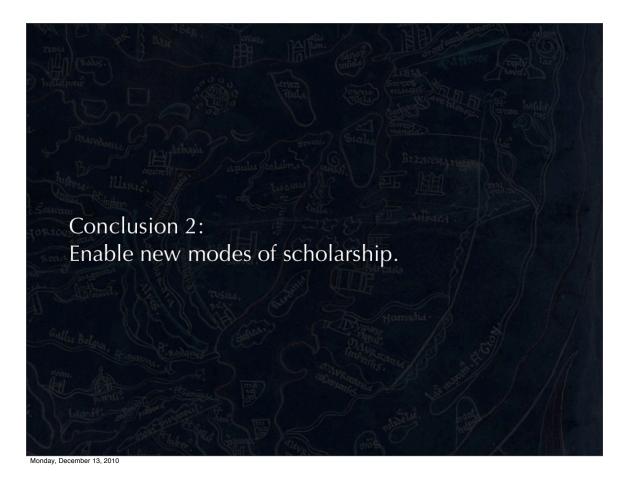
B.8: demonstration of interoperability proof of concept (con't); link from Parker resource opens folio page called from Parker in Phase 1 DM for editing. After editing, Parker interface dynamically shows that edits in *DM* resource now exist and may be viewed.



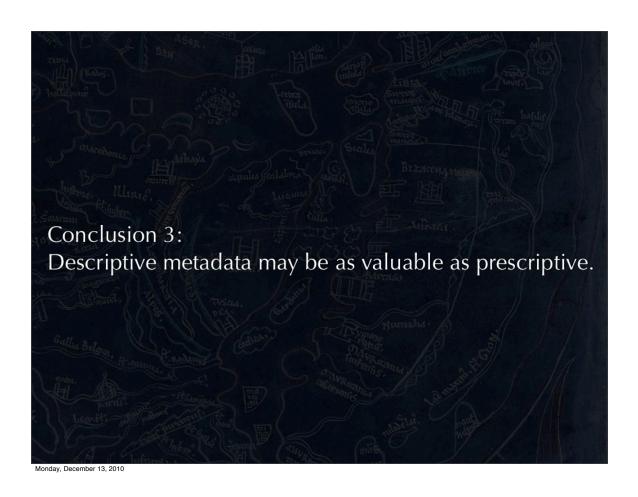


B.9: phase 2 annotation philosophy; scholarly primitives and de-taxonomics





B.10 : phase 2 annotation philosophy (con't); Gmail example of "loose" data management driven by search, not taxonomy



B.11: remembering the simplest formula for metadata